DOCUMENT RESUME

ED 410 300 TM 027 179

AUTHOR	Flowers, Claudia P.; And Others
TITLE	The Relationship between Polytomous DFIT and Other Polytomous DIF Procedures.
PUB DATE	Mar 97
NOTE	24p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, March 25-27, 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Chi Square; Computer Simulation; *Item Bias; Item Response Theory; *Sample Size; Tables (Data); *Test Items
IDENTIFIERS	Dimensionality (Tests); Item Bias Detection; *Polytomous Items; SIBTEST (Computer Program); Type I Errors

ABSTRACT

An item response theory-based parametric procedure proposed by N. S. Raju, W. J. van der Linden, and P. F. Fleer (1995) known as differential functioning of items and tests (DFIT) can be used with unidimensional and multidimensional data with dichotomous or polytomous scoring. This study describes the polytomous DFIT framework and evaluates and compares its performance to that of the extension of the SIBTEST procedure developed by R. Shealy and W. Stout (1993) and extended Lord's chi-square. Using simulated data, the effects of sample size (500 and 1,000 examinees), focal group distribution (N(0,1) and N(-1,1)) number of differential item functioning (DIF) items (0%, 10%, and 20%), magnitude of DIF, and the value of a-parameter were evaluated. Overall, the DFIT framework performed well. Type I error rates were affected by the number of DIF items, magnitude of DIF, and the value of the a-parameters. The DIF detection rates were affected by all the factors in the study. Future directions for research are discussed. (Contains 5 tables and 16 references.) (Author/SLD)

THE RELATIONSHIP BETWEEN POLYTOMOUS DFIT AND OTHER POLYTOMOUS DIF PROCEDURES

Claudia P. Flowers
The University of North Carolina at Charlotte

T. C. Oshima
Georgia State University

Nambury S. Raju
Illinois Institute of Technology

Paper presented at NCME Annual Meeting, Chicago, March 1997

Running Head: Polytomous DFIT

ABSTRACT

An IRT-based, parametric procedure proposed by Raju, van der Linden, and Fleer (1995), known as differential functioning of items and tests (DFIT), can be used with unidimensional and multidimensional data with dichotomous and/or polytomous scoring. The purpose of this study is to describe the polytomous DFIT framework and evaluate and compare its performance to the extension of Shealy and Stout's (1993) SIBTEST and the extended Lord's chi-square. Using simulated data, the effects of sample size (500 and 1000 examinees), focal group distribution (N(0,1) and N(-1,1)), number of DIF items (0%, 10%, and 20%), magnitude of DIF, and value of a-parameter were evaluated. Overall, the DFIT framework performed well. Type I error rates were affected by the number of DIF items, magnitude of DIF, and the value of the a-parameters. The DIF detection rates were affected by all the factors in the study. Future directions for research are discussed.

THE RELATIONSHIP BETWEEN POLYTOMOUS DFIT AND OTHER POLYTOMOUS DIF PROCEDURES

The increased use of polytomously-scored items on tests has stimulated interest in polytomous differential item functioning (DIF) procedures. Many polytomous DIF procedures have been proposed over the last four years (e.g., combined *t* tests (Welch & Hoover, 1993); an extension of Shealy and Stout's (1993) SIBTEST procedure (Chang, Mazzeo, & Roussos, 1996; Mazzeo & Chang, 1994); an application of logistic discrimination function analysis (Miller & Spray, 1993); logistic regression approaches (Rogers & Swaminathan, 1994); extensions of Lord's chi-square, signed area, and unsigned area (Cohen, A.S., Kim, S., & Baker, F.B., 1993)).

An IRT-based, parametric procedure proposed by Raju, van der Linden, and Fleer (1995), known as differential functioning of items and tests (DFIT), can be used with unidimensional and multidimensional data with dichotomous and/or polytomous scoring. The purpose of this study is to describe the polytomous DFIT framework and compare its performance to the extension of Shealy and Stout's (1993) SIBTEST (Chang, Mazzeo, & Roussos, 1996; Mazzeo & Chang, 1994) and the extended Lord's chi-square (Cohen, A.S., Kim, S., & Baker, F.B., 1993). Both DFIT and SIBTEST have differential test functioning (DTF) procedures, but they will not be examined in this study. The first section of this study provides a definition of DIF. The second section describes Raju et al. (1995) polytomous-DFIT framework. A brief explanation is provided of the extensions of SIBTEST and Lord's chi-square. The third section presents the results of a simulation. The final section summarizes the finding and suggests direction for future research.

Definition of DIF

Chang and Mazzeo (1994) demonstrated that for the graded response model (Samejima, 1969), partial credit model (Masters, 1982), and generalized partial credit model (Muraki, 1992), if two items have the same item response functions (IRF) then they must have the same number of scoring categories and same item category response functions (ICRF). An IRF for a polytomous item can be expressed as

$$E_g[Y|\theta] = \Sigma k P_{k,g}(\theta) \tag{1}$$

where $E_R[Y|\theta]$ is the item expected score (Y) for group g at a given θ level, $P(\theta)$ is the ICRF for group g with a category score of k. In other words, the expected item score is a weighted sum of the ICRFs. The null DIF hypothesis would be

$$E_R[Y|\theta] = E_F[Y|\theta] \tag{2}$$

where $E_R[Y|\theta]$ is the item expected score for an examinee in the reference group with a given θ and $E_F[Y|\theta]$ is the item expected score for an examinee in the focal group with a given θ.

Polytomous DIF Procedures

Potenza and Dorans (1994) proposed a framework for classifying polytomous DIF procedures. First, they distinguished between procedures that use an observed score as a matching variable and procedures that match groups in terms of an estimate of a latent variable. Secondly, they distinguish between parametric approaches that assume a parametric functional form for the item response function (IRF) and procedures that do not make such assumptions (i.e., nonparametric approaches).

Using this classification system, all the DIF procedures in this study use an estimate of the latent variable as a matching variable. The only difference between the procedures is that DFIT and Lord's chi-square are parametric approaches (i.e., require IRT ability and item parameter estimations) and the extension of SIBTEST is a nonparametric approach.

The Polytomous DFIT Framework

Raju et al. (1995) suggest that for polytomously-scored data an expected score (ES_{si}) for item i can be computed for examinee s as

$$ES_{si} = \sum_{k=1}^{m} P_{ik}(\Theta_s) X_{ik} \tag{3}$$

where X_{ik} is the score or weight for category k; m is the number of categories; and P_{ik} is the probability of responding to category k (similar to Equation 1). Summing the expected item scores across a test will result in the true test score function for each examinee as

$$T_s = \sum_{i=1}^{n} ES_{si} \tag{4}$$

where n is the number of items in the test. The null hypothesis for DTF would be

$$T_R = T_F \tag{5}$$

where T_R and T_F are the true test scores for the reference and focal group examinees with the same θ, respectively.

The difference between the dichotomous and polytomous DFIT framework is the calculation of the item true score. The item true score must accommodate the multiple categories in the polytomous model (see Equation 3). Once the true item and test scores are known, the DFIT framework for the polytomous framework is identical to the DFIT framework for the dichotomous case.

According to Raju et al. (1995), a measure of DTF at the examinee level may be defined as

$$D_s^2 = (T_{sF} - T_{sR})^2 \; . \tag{6}$$

DTF across the focal group examinees may be defined as

$$DTF = \underset{F}{\varepsilon} D_s^2 = \underset{F}{\varepsilon} (T_{sF} - T_{sR})^2 \tag{7}$$

or, equivalently,

$$DTF = \int_{\theta} D_s^2 f_F(\theta)\, d\theta \tag{8}$$

where $f_F(\theta)$ is the density function of θ for the focal group. Also,

$$DTF = \sigma_D^2 + (\mu_{TF} - \mu_{TR})^2 = \sigma_D^2 + \mu_D^2 \qquad (9)$$

where μ_{TF} is the mean true score for the focal group examinees; μ_{TR} is the mean true score for the same examinees as if they were members of the reference group; and σ_D^2 is the variance of D.

Differential functioning at the item level can be derived from Equation 7. If

$$d_{si} = ES_{siF} - ES_{siR} \qquad (10)$$

then

$$DTF = \epsilon[(\sum_{i=1}^{n} d_{si})^2] \qquad (11)$$

where n is the number of items in a test. This can be rewritten as

$$DTF = \sum_{i=1}^{n} [Cov(d_i, D) + \mu_{d_i}\mu_D] \qquad (12)$$

where $Cov(d_i,D)$ is the covariance of the difference in expected item scores (d_i) and the difference in true scores (D), and μ_{di} and μ_D are the means of d_{is} and D_s, respectively. In this case DIF can be written as

$$DIF_i = Cov(d_i, D) + \mu_{d_i}\mu_D. \qquad (13)$$

Raju et al. (1995) refer to this DIF as compensatory DIF (CDIF). If DIF in Equation 13 was expressed as CDIF, then Equation 12 can be rewritten as

$$DTF = \sum_{i=1}^{n} CDIF_i \ . \tag{14}$$

The additive nature of DTF allows for possible cancellation at the test level. This occurs when one item displays DIF in favor of one group and another item displays DIF in favor of the other group. This combination of DIF items will have a canceling effect on the overall DTF. The sum of the CDIF indices reflects the net directionality.

Raju et al. (1995) proposed a second index, named noncompensatory DIF (NCDIF) that assumes that all items other than the one under study are free from differential functioning. In the dichotomous case, NCDIF is closely related to other existing DIF indices such as Lord's chi-square and the unsigned area (Raju et al., 1995). If all other items are DIF free, then $d_j = 0$ for all $j \neq$ i where i is the item being studied and Equation 13 can be rewritten as

$$NCDIF_i = \sigma_{d_i}^{\ 2} + \mu_{d_i}^{\ 2} . \tag{15}$$

Raju et al. (1995) noted that items having significant NCDIF do not necessarily have significant CDIF in the sense of contributing significantly to DTF. For example, if one item favors the reference group and another item favors the focal group, significant NCDIF occurs for both items even though the two CDIF indices may not be significant because of their canceling effect at the test level. This could lead to a greater number of significant NCDIF items than CDIF items.

Statistical significance testing can be performed but these tests have been shown to be overly sensitive for large sample sizes (Fleer, 1993). Fleer suggested empirically establishing a critical (cutoff) value for NCDIF. This critical value was determined from a Monte Carlo study of non-DIF items.

Extension of SIBTEST

Chang, Mazzeo, and Roussos (1996) provides a detailed explanation of the extension of the SIBTEST. The amount of DIF is measured by

$$B_0(\theta) \equiv E_R[Y|\theta] - E_F[Y|\theta] \qquad (16)$$

Shealy and Stout (1993) provide a global index of DIF as

$$\beta = \int B_0(\theta) f_F(\theta) d\theta. \qquad (17)$$

This is interpreted as the expected amount of DIF experienced by a randomly selected focal group examinee.

Two minor modifications to the original SIBTEST are needed to accommodate polytomous data: (1) replacement of n (i.e., number of items) in the SIBTEST test statistic (Shealy & Stout, 1993) with n_h (maximum test score) and (2) modify the matching test reliability estimates used by Shealy and Stout in their regression correction, substituting Cronbach's alpha for KR 20 (Chang, Mazzeo, & Roussos, 1996).

Lord's Chi-Square

Lord's chi-square simultaneously tests the difference between the a and b-parameters for each group. In the dichotomous case, a vector of differences between the parameters is calculated. A similar method is used in the polytomous case with the exception that a larger number of elements would be included in the vector because of the multiple b-parameters in the polytomous model. Cohen, Kim, and Baker (1993) offer the following extension of the polytomous Lord's chi-square:

$$\xi_j = [a_{jR} - a_{jF},\ b_{j1F} - b_{j1R},\ \ldots\ ,\ b_{j(m-1)F} - b_{j(m-1)R}]' \tag{18}$$

where j is the item under study, and m is the number of categories. Then

$$\chi_j^2 = \xi_j' \Sigma_j^{-1} \xi_j \tag{19}$$

where Σ_j is the variance-covariance matrix of the difference between item parameters. There are m degrees of freedom for this extension of Lord's chi-square.

Simulation Study

To evaluate the performance of the DFIT framework, a simulations study was conducted. The DIF procedure (NCDIF) was the only DFIT index evaluated in this study. The results from DFIT were compared to the extensions of SIBTEST and Lord's chi-square. The simulation represented a 20-item test with all items having five-category responses. Item response data were generated using the graded response model. The reference group item parameters are contained in Table 1.

Insert Table 1 about here

Factors Examined

DIF and null DIF modeling. DIF was modeled by adding a constant to the b-parameters of the focal group (i.e., $b_{ikF} = b_{ikR} + C_{ik}$ where i is the item and k is the item boundary). For the null condition, C_{ik} was equal to zero. For the DIF conditions, two magnitudes of DIF were embedded: (1) $C_{ik} = .10$ and (2) $C_{ik} = .25$. The number of DIF items was also varied across test conditions: (1) two DIF items and (2) four DIF items. For the two-DIF items conditions, items 4 and 17 were embedded with DIF. For the four-items DIF conditions, items 1, 4, 10, and 17 were embedded with DIF. It should be noted that the b-parameters across these items are fixed (i.e., have the same value) but the a-parameters vary across items (i.e., item 1 = .55, item 4 = .75, item 10 = 1.00, and item 17 = 1.36).

Other factors. Two sample sizes were simulated. In one condition, the focal and reference groups each had 500 examinees, and for the other condition, the focal and reference groups each had 1000 examinees. Additionally, two focal group ability distributions were simulated: N(0,1) and N(-1,1). The reference group ability distribution was N(0,1) for all simulation conditions. Simulation under each factor combination was iterated 100 times. The nominal alpha used for detecting DIF was 0.05.

Calculation of DIF Indices

Both DFIT and Lord's chi-square calculations required item and ability estimations as well as an equating procedure. All item and ability parameters were estimated using the computer program PARSCALE 2 (Muraki & Bock, 1993). The maximum marginal likelihood procedure and EM algorithm were used to estimate the item parameters. Default values were used for all estimations. Estimation of underlying abilities were made using Bayesian EAP procedure which incorporates normal priors. The estimation of equating coefficients was made by means of Baker's modified test characteristic curve method as implemented in the EQUATE 2.0 computer program (Baker, 1993). In this study, all parameter estimates for the reference group were equated to the underlying metric of the focal group. A Fortran program written by Raju (1995) was used to calculate the DFIT indices. A Fortran program written by Kim (1993) was used to calculate Lord's chi-square.

Recall that DFIT statistical test are overly sensitive to large sample sizes. Critical (cutoff) values were established independently of the current study by simulating 2000 nonDIF items and noting the value at the 95th percentile. The critical values used in this study were .011 for the 500 examinee condition and .05 for the 1000 examinee condition. This would be equivalent to a nominal alpha of .05.

A computer program, PSIBTEST, written by Roussos, Shealy, and Chang (1993) was used to detect DIF for the extension of SIBTEST. This program does not require the estimation of item parameters or equating.

Results

The results will be divided into two sections: Type I error rate and DIF detection rate. Five effects are discussed in each section: (1) number of examinees (500 and 1000); (2) focal group distribution (N(0,1) and N(-1,1)); (3) number of DIF items (0, 2, and 4); (4) magnitude of DIF (.10 and .25); (5) item discrimination (.55, .75, 1.00, and 1.36).

Comparisons of the effectiveness between the DIF indices should not be made. As mentioned previously, critical values for detecting DIF in DFIT were established using empirical data. The performance of SIBTEST and Lord's chi-square would improve if this method was used in establishing critical values for these DIF procedures. This study focuses on the effects of the factors being manipulated on the performance of the DIF indices.

Type I Error Rate

Table 2 contains the average Type I error rate for all conditions.

Insert Table 2 about here

The effects of sample size can be examined by looking across conditions in Table 2. The sample size had little effect on the NCDIF error rate. Most conditions were close to the alpha level (i.e., .05). The only exception was in the condition with the greatest number of DIF items (four items) and the greatest magnitude of DIF (.25). In this condition the Type I error rate increased (ranging from .09 to .14) with the 1000 examinee conditions having the greatest error rate. Sample size had little effect on SIBTEST when the reference and focal group had equal ability distributions (Focal:N(0,1)) but the error rate almost doubled in the 1000 examinee condition with a focal group distribution of N(-1,1). Lord's chi-square had higher rate than expected in all conditions with the 1000 examinee condition usually having the highest error rate.

NCDIF was not affected by the focal group distribution. The Type I error rates were almost identical in all conditions. Focal group distribution had the most noticeable effect on SIBTEST. In all canteens the Type I error rate increased in the Focal:N(-1,1) condition. This is consistent with previous studies which showed that there is an over-regression-correction (Chang, Mazzeo, & Roussos, 1996) when focal and reference group distributions are not equivalent. Lord's chi-square had a slight increase in the error rate for the Focal:N(-1,1) in almost all conditions.

All indices showed an increase error rate as the number of DIF items and the magnitude of the DIF increased. As expected, the condition with the most DIF items and the greatest magnitude had the largest Type I error rate.

Table 3 contains the Type I error rate for the studied items (i.e., items 1, 4, 10, and 17) in the null condition. The value of the a-parameters had an effect on the Type I error rate. For NCDIF, smaller a-parameters resulted in higher Type I error rates. No trend was noted with SIBTEST or Lord's chi-square.

Insert Table 3 about here

DIF Detection Rate

Table 4 contains the average DIF detection rate for all conditions.

Insert Table 4 about here

For all the DIF indices, in almost all conditions, the detection rate was consistently higher in the 1000 examinee conditions than the 500 examinee conditions. NCDIF and SIBTEST had consistently lower detection rates in the Focal:N(-1,1) than the Focal:N(0,1). This pattern was not noted for Lord's chi-square. For all indices, the average detection rate decreased as the number of

DIF items increased. As expected, all the indices had a much higher detection rate for the .25 condition than the .10 condition.

Table 5 contains the detection rate by the studied items. When the magnitude of DIF was .25, high discriminating items had a better detection rate for NCDIF. This trend was not noted in the .10 magnitude condition. Both SIBTEST and Lord's chi-square had higher detection rates as the item discrimination increased in all conditions.

Insert Table 5 about here

Summary

This study supports the validity of the DFIT frameworks in detecting DIF in polytomous data. The Type I error rates were close to nominal alpha level except when the number of DIF items (i.e., 20% of the test) and the magnitude of DIF was highest. This is typically true for most DIF procedures. The only other factor that affected the Type I error rate was the value of the a-parameters; that is, lower a-parameters had higher Type I error rates. The DIF detection rate was affected by all the factors in this study. DFIT detects DIF items better for; (1) larger sample sizes, (2) equivalent focal and reference group distributions, (3) fewer DIF items in a test, (4) greater magnitude of DIF, and (5) larger a-parameter values.

The results of this study encourage further research of the DFIT framework. First, a statistical test that is not as sensitive to sample size needs to be developed. Second, the DTF procedure (i.e., CDIF) needs to be investigated. Third, the DFIT framework needs to be applied to a mixed test format (i.e., dichotomous and polytomous items). Fourth, the ability to detect different types of DIF (i.e., uniform and nonuniform) needs to be examined.

References

Baker, F. B. (1993). EQUATE2: Computer program for equating two metrics in item response theory [computer program]. Madison:University of Wisconsin, Laboratory of Experimental Design.

Chang, H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. Journal of Educational Measurement, 33, 333-353.

Chang, H., & Mazzeo, J. (1994). The unique correspondence of the item response function and the item category response functions in polytomously scored item response models. Psychometrika, 59, 391-404.

Cohen, A.S., Kim, S., & Baker, F.B. (1993). Detection of differential item functioning in the graded response model. Applied Psychological Measurement, 17, 335-350.

Fleer, P. F. (1993). A Monte Carlo assessment of a new measure of item and test bias. Unpublished PhD dissertation, Illinois Institute of Technology.

Masters, G.N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Mazzeo, J., & Chang, H. (1994). Detecting DIF for polytomously scored items: Progress in adaptation of Shealy-Stout's SIBTEST procedure. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Miller, T.R., & Spray, J.A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. Journal of Educational Measurement, 30, 107-122.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16, 159-176.

Muraki, E., & Bock, R.D. (1993). PARSCALE2: IRT based test scoring and item analysis for graded open-ended exercises and performance task [computer program]. Chicago: Scientific Software International.

Potenza, M.T., & Dorans, N.J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. Applied Psychological Measurement, 19, 23-27.

Raju, N. (1995). DFITPU: A Fortran program for calculating DIF/DTF. [computer program]. Georgia Institute of Technology.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measure of differential functioning of items and tests. Applied Psychological Measurement, 19, 353-368.

Rogers, H.J., & Swaminathan (1994). Logistic regression procedures for detecting DIF in nondichotomous item responses. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Roussos, L., Shealy, R., & Chang, H. (1993). A Fortran program for calculating the extension of SIBTEST [computer program]. University of Illinois at Urbana-Champaign.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph, No. 17.

Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. Psychometrika, 58, 159-194.

Welch, C., & Hoover, H.D. (1993). Procedures for extending item bias detection techniques to polytomously scored items. Applied Measurement in Education, 6, 1-19.

Table 1

Reference Group Item Parameters Used in Simulation Study

Item Number	a_i	b_{i1}	b_{i2}	b_{i3}	b_{i4}
1	0.55	-1.80	-0.60	0.60	1.80
2	0.73	-2.32	-1.12	0.08	1.28
3	0.73	-1.80	-0.60	0.60	1.80
4	0.73	-1.80	-0.60	0.60	1.80
5	0.73	-1.28	-0.08	1.12	2.32
6	1.00	-2.78	-1.58	-0.38	0.82
7	1.00	-2.32	-1.12	0.08	1.28
8	1.00	-2.32	-1.12	0.08	1.28
9	1.00	-1.80	-0.60	0.60	1.80
10	1.00	-1.80	-0.60	0.60	1.80
11	1.00	-1.80	-0.60	0.60	1.80
12	1.00	-1.80	-0.60	0.60	1.80
13	1.00	-1.28	-0.08	1.12	2.32
14	1.00	-1.28	-0.08	1.12	2.32
15	1.00	-0.82	0.38	1.58	2.78
16	1.36	-2.32	-1.12	0.08	1.28
17	1.36	-1.80	-0.60	0.60	1.80
18	1.36	-1.80	-0.60	0.60	1.80
19	1.36	-1.28	-0.08	1.12	2.32
20	1.80	-1.80	-0.60	0.60	1.80

Table 2

Type I Error Rate ($\alpha = .05$) for All Conditions

	NCDIF		SIBTEST		Lord's χ^2	
	Number of Examinees		Number of Examinees		Number of Examinees	
Condition	500	1000	500	1000	500	1000
Null						
Focal:N(0,1)	.06	.04	.05	.05	.13	.08
Focal:N(-1,1)	.05	.04	.08	.15	.12	.12
Constant .10						
2 DIF Items						
Focal:N(0,1)	.05	.06	.05	.06	.12	.13
Focal:N(-1,1)	.05	.06	.09	.19	.13	.15
4 DIF Items						
Focal:N(0,1)	.06	.06	.06	.06	.13	.15
Focal:N(-1,1)	.06	.07	.12	.22	.16	.17
Constant .25						
2 DIF Items						
Focal:N(0,1)	.07	.07	.07	.08	.14	.17
Focal:N(-1,1)	.07	.06	.12	.25	.17	.17
4 DIF Items						
Focal:N(0,1)	.09	.14	.12	.15	.22	.30
Focal:N(-1,1)	.10	.12	.17	.37	.24	.31

Table 3

Type I Error Rate ($\alpha = .05$) for Studied Item in the Null Condition

		NCDIF		SIBTEST		Lord's χ^2	
		Number of Examinees		Number of Examinees		Number of Examinees	
Condition	a-parameter	500	1000	500	1000	500	1000
Null							
Focal:N(0,1)							
Item 1	.55	.15	.16	.05	.03	.10	.07
Item 4	.73	.09	.08	.06	.06	.09	.09
Item 10	1.00	.08	.02	.02	.03	.15	.07
Item 17	1.36	.04	.00	.04	.05	.15	.09
Focal:N(-1,1)							
Item 1	.55	.12	.11	.05	.10	.07	.05
Item 4	.73	.14	.10	.08	.11	.16	.17
Item 10	1.00	.02	.02	.08	.15	.11	.13
Item 17	1.36	.04	.03	.12	.12	.12	.14

Table 4

Average DIF Detection Rate for Each Condition ($\alpha = .05$)

	NCDIF		SIBTEST		Lord's χ^2	
	Number of Examinees		Number of Examinees		Number of Examinees	
Condition	500	1000	500	1000	500	1000
Constant .10						
2 DIF Items						
Focal:N(0,1)	.30	.44	.30	.38	.43	.63
Focal:N(-1,1)	.22	.42	.15	.15	.40	.62
4 DIF Items						
Focal:N(0,1)	.25	.40	.21	.28	.34	.52
Focal:N(-1,1)	.20	.30	.12	.10	.34	.47
Constant .25						
2 DIF Items						
Focal:N(0,1)	.91	.98	.85	.94	.91	.98
Focal:N(-1,1)	.87	.98	.72	.66	.93	.98
4 DIF Items						
Focal:N(0,1)	.78	.94	.72	.80	.79	.92
Focal:N(-1,1)	.73	.94	.54	.48	.79	.94

Table 5

Detection Rate by DIF Item ($\alpha = .05$)

		NCDIF		SIBTEST		Lord's χ^2	
		Number of Examinees		Number of Examinees		Number of Examinees	
Condition		500	1000	500	1000	500	1000
Constant .10							
2 DIF Items							
Focal:N(0,1)	Item 4	.35	.45	.19	.22	.32	.46
	Item 17	.24	.42	.40	.53	.54	.79
Focal:N(-1,1)	Item 4	.25	.38	.10	.13	.29	.43
	Item 17	.18	.46	.19	.16	.50	.80
4 DIF Items							
Focal:N(0,1)	Item 1	.24	.41	.10	.16	.14	.34
	Item 4	.29	.42	.14	.13	.30	.44
	Item 10	.24	.41	.24	.34	.39	.54
	Item 17	.22	.34	.35	.47	.51	.74
Focal:N(-1,1)	Item 1	.16	.27	.06	.08	.16	.21
	Item 4	.26	.26	.09	.10	.30	.32
	Item 10	.20	.39	.16	.09	.41	.58
	Item 17	.16	.29	.15	.12	.50	.76
Constant .25							
2 DIF Items							
Focal:N(0,1)	Item 4	.82	.96	.71	.87	.81	.95
	Item 17	1.00	1.00	.98	1.00	1.00	1.00
Focal:N(-1,1)	Item 4	.81	.96	.52	.44	.87	.95
	Item 17	.93	1.00	.92	.88	.98	1.00
4 DIF Items							
Focal:N(0,1)	Item 1	.57	.82	.36	.48	.48	.71
	Item 4	.72	.95	.64	.74	.72	.95
	Item 10	.92	1.00	.91	.97	.96	1.00
	Item 17	.89	1.00	.96	1.00	.98	1.00
Focal:N(-1,1)	Item 1	.60	.85	.29	.24	.55	.81
	Item 4	.65	.92	.34	.40	.67	.94
	Item 10	.84	.99	.71	.60	.93	.99
	Item 17	.83	.99	.81	.68	.98	1.00

AERA April 8-12, 1996

TM027179

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

ERIC®

I. DOCUMENT IDENTIFICATION:

Title: THE RELATIONSHIP BETWEEN POLYTOMOUS DFIT AND OTHER POLYTOMOUS DIF PROCEDURES	
Author(s): FLOWERS, C.P., OSHIMA, T.C., & RAJU, N.S.	
Corporate Source:	Publication Date: 1997

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[✓] **Sample sticker to be affixed to document**

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here
Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Claudia Flowers	Position: ASSISTANT PROFESSOR
Printed Name: CLAUDIA FLOWERS	Organization: UNC CHARLOTTE
Address: UNC CHARLOTTE 9201 UNIV. CITY BLVD CHARLOTTE, NC 28223-0001	Telephone Number: (704) 547-4545
	Date: 5-13-97